

Kutatói tájékoztató

2022. december

TARTALOM

| | |
|--|----|
| I. A KUTATÓSZOBA HASZNÁLATA | 3 |
| 1. Bevezetés - a kutatói tájékoztató célja | 3 |
| 2. A Kutatószoba működése - használata | 3 |
| 2.1. Nyitvatartás, időpontfoglalás | 3 |
| 2.2. A Kutatószobába történő belépés előtt | 4 |
| 2.3. Kutatószobai munka | 4 |
| II. KUTATÁSI EREDMÉNYEK..... | 7 |
| 1. Bevezetés | 7 |
| 2. A Kutatás dokumentációja | 8 |
| 2.1 Kutatói munka során elvárt könyvtárstruktúra | 8 |
| 2.2 Segédtablázatok előállítása | 10 |
| 2.3 Dokumentáció a programfájlokban | 10 |
| 2.4 Nem programkóddal előállított kutatási eredmények dokumentálása | 11 |
| 2.5 A kutatási eredmények formátuma..... | 11 |
| 2.6 Dokumentációs űrlap | 12 |
| 3. Kutatási eredmények utólagos adatvédelmi szempontú ellenőrzése (output checking) | 12 |
| 3.1 Kutatási eredményekre vonatkozó adatvédelmi szabályok | 12 |
| 3.2 Az egyes kutatási eredménytípusok esetén figyelembe vett adatvédelmi szabályok | 14 |
| 3.3. Javaslatok | 14 |
| 4. Külső állományok | 15 |
| 1. Melléklet: Példák programfájlok dokumentálásához | 16 |
| 2. Melléklet: Példa értékösszeg táblák adatvédelmi ellenőrzésére | 17 |
| 3. Melléklet: Példa táblázatos adat átsztruktúrálására | 20 |
| 4. Melléklet: Dokumentációs űrlap | 21 |

I. A KUTATÓSZOBA HASZNÁLATA

1. BEVEZETÉS - A KUTATÓI TÁJÉKOZTATÓ CÉLJA

A Kutatószoba a KSH belső rendszereitől elkülönítetten működő biztonságos, kamerás megfigyeléssel biztosított környezet, amely kizárólag tudományos célból biztosít hozzáférést közvetlen azonosításra alkalmatlan adatállományokhoz az egyedi statisztikai adatok kiemelt szintű védelme, illetve az adatvédelemmel kapcsolatos jogszabályok szigorú betartása mellett.

Jelen dokumentum célja, hogy a kutatók részletes információt kapjanak a KSH Kutatószobájának működéséről, valamint azokról a szabályokról, amelyeket - a hatékony adatvédelmi ellenőrzés biztosítása, ezáltal a kutatási eredmények gyorsabb kiadása érdekében – követniük kell a kutatási eredmények (outputok) összeállítása során.

2. A KUTATÓSZOBA MŰKÖDÉSE - HASZNÁLATA

2.1. NYITVATARTÁS, IDŐPONTFOGLALÁS

A KSH Kutatószobát munkanapokon hétfőtől csütörtökig 09:00-16:00-ig, pénteken 09:00-13:00-ig lehet használni.

- **Időpontfoglalás nélkül tilos a Kutatószobába történő belépés!**
- **A nyitvatartási időn túl tilos a Kutatószobában tartózkodni!**

A tervezett első kutatási alkalmat a Kutatónak a kutatooszoba@ksh.hu email címen szükséges egyeztetnie a Kutatószobai koordinátorral.

Az egyeztetett időpontban, miután a Kutató a KSH fő portáján (1024 Budapest, Keleti Károly u. 5-7.) megkapja a belépéshez szükséges belépőkártyát, a Kutatószobai koordinátor bekíséri őt a Kutatószoba előterébe és tájékoztatja a **Kutatószobába történő belépést megelőző alapvető szabályokról**.

Az első alkalmat követően már nem szükséges a Kutatószobai Koordinátorral egyeztetni a további a kutatási alkalmakról, azonban a Kutatószoba használatát megelőzően **időpontot kell foglalni a**

[KSH foglalási rendszerében](#)

A foglalási rendszerben jelzett kutatási napon a Kutató a KSH fő portáján megkapja a belépéshez szükséges belépőkártyát és önállóan veszi igénybe a Kutatószoba szolgáltatásait.

2.2. A KUTATÓSZOBÁBA TÖRTÉNŐ BELÉPÉS ELŐTT

- **A Kutatószobába semmilyen személyes tárgy (táska, kabát) nem vihető be!**

Belépés előtt ezért a személyes tárgyakat az erre a célra kialakított kulccsal zárható szekrényekben kell elhelyezni.

A kutatók számára tilos továbbá:

- kommunikációs, információ rögzítésére alkalmas eszköz (**telefon, laptop**) bevitelle, illetve bármilyen célra történő használata,
- bármilyen **papíralapú dokumentum bevitelle, használata** (könyv, jegyzet stb.),
- **érvényes kutatási projekt nélkül történő belépés** a Kutatószobába.

2.3. KUTATÓSZOBAI MUNKA

A Koordinátor – az első kutatási napon – átadja a kutatói felületre való belépéshez szükséges azonosítót és jelszót.

- **A kapott jelszót a Kutató az első bejelentkezéskor köteles megváltoztatni!**

Mappaszerkezet

A Koordinátor megmutatja a kutató felületét, a mappaszerkezetet és megnézik, hogy a kért adatállományok teljes köre kihelyezésre került-e.

- A Kutató rendelkezésére bocsátott adatok, valamint a Kutató által bevinni kívánt fájlok a kutatói felület **Bejövő mappájában** kerülnek elhelyezésre. A Bejövő mappában a Kutatónak csupán olvasási jogosultsága van, ezen a felületen dolgozni nem tud.
- Az elemzéseket, valamint ezek eredményeit a **Kimenő mappában** a Kutatási eredmények fejezetben leírt mappastruktúra szerint tudja elhelyezni.

Jegyzet készítése

Jegyzet készítésére csak a Kutatószobában elhelyezett **fehér táblákon** (a szintén kihelyezett filctollakkal), valamint a kutatói felületen elérhető alkalmazásokkal, **elektronikus formában (pl. Word, Notepad) van lehetőség**. Az utóbbi, elektronikus dokumentum ki is kérhető a Kimenő felület megfelelő mappájába másolva.

- **A jegyzet csak elektronikus formában, adatvédelmi ellenőrzést követően vihető ki a Kutatószobából!**

Kutatási eredmények, valamint az **elektronikus formában** készített jegyzet kikérését a Kutatószobából való távozás után szükséges jelezni a kutatoszoba@ksh.hu e-mail címen keresztül.

A kutatói munka során a **Kutatószobában tilos:**

- **bármilyen eszköz kliensgépekhez történő csatlakoztatása** vagy arra utaló próbálkozás,
- **adatállományok** bármilyen **külső adathordozóra történő másolása**, vagy arra utaló próbálkozás,
- **kamera szándékos eltakarása, rongálása,**
- a Kutatószobában található eszközökben kárt okozni,

Kutatás dokumentációja

A Kutatói tájékoztató II.2. fejezetében leírtak alapján szükséges elkészíteni a: kutatói munka során elvárt könyvtárstruktúrát, a kutatási eredményekhez tartozó segédtablázatokat, a kutatási eredmények programkóddal vagy programkód nélküli **szöveges** dokumentálását. Ebben az anyagban kerül felsorolásra, hogy milyen kiterjesztésű fájlok adhatók be adatvédelmi ellenőrzésre. Továbbá az outputok kikéréséhez szükséges kitölteni az úgynevezett „Dokumentációs űrlapot”. Addig nem kezdődik el a beadott outputok ellenőrzése, míg az űrlap hiányos vagy rosszul kitöltött.

Kutatói támogatás

A Kutatószobai Koordinátor az első kutatási napon tájékoztatást nyújt a kutató számára a segítségkérési lehetőségekről.

Amennyiben a kutatás során bármilyen kérdés, probléma – pl. elfelejtett jelszó – merülne fel, a kutató (a Kutatószobában is kihelyezett) alábbi telefonszámokon tud kapcsolatba lépni a Kutatószobai koordinátorral:

- 6621
- 6924

Incidensek, szankciók

Amennyiben a Kutatószobát használó kutató megsérti a Kutatószoba használatának szabályait, a kötelezettségszegés jellegétől függően a KSH az alábbi szankciókat jogosult alkalmazni:

- Amennyiben a **kötelezettségszegés adminisztratív jellegű** és nem veszélyeztette súlyosan a kutatószobai adatok védelmét – például a Kutató nem regisztrált a foglalási rendszerben – **a KSH a kutatót írásban figyelmezteti.** Három egymást követő figyelmeztetés után – amely figyelmeztetéseknek nem kell ugyanazon kutatáshoz kapcsolódniuk – a KSH jogosult a kutatói hozzáférést felfüggeszteni és a Kutatót a KSH-ban működő Kutatószoba használatától 2 évre eltiltani.
- Ha a **kötelezettségszegés biztonsági incidenst idézett elő**, amely ugyanakkor nem veszélyeztette jelentősen a kutatószobai adatok védelmét, a **KSH a kötelezettségszegést írásban rögzíti. A második biztonsági incidens előidézése esetén a KSH jogosult a Kutatót a Kutatószoba használatától 5 évre eltiltani.** Ilyen biztonsági incidensnek minősül különösen:

- kommunikációs, rögzítésre alkalmas eszköz bevitele
- bármilyen írott dokumentum bevitele, kivitele (könyv, jegyzet stb.)
- **Ha a kötelezettségszegés olyan biztonsági incidenst idézett elő, amely jelentősen veszélyeztette a kutatószobai adatok védelmét, a kötelezettségszegésről szóló jegyzőkönyv felvétele mellett a KSH a kutatót azonnali hatállyal, végleg eltilthatja a KSH-ban működő Kutatószoba használatától.** Ilyen biztonsági incidensnek minősül különösen:
 - kommunikációs, információ rögzítésére és/vagy továbbítására alkalmas eszköz bármilyen célra történő használata
 - saját jegyzetek kivitele
 - jegyzet kutatószobából történő elvitele
 - internet, e-mail használat
 - bármilyen eszköz zéró klienshez történő csatlakoztatása
 - adatállományok bármilyen külső adathordozóra történő másolása, vagy arra utaló próbálkozása
 - kamera szándékos eltakarása, rongálása
 - érvényes kutatási projekt nélkül történő belépés a Kutatószobába

II. KUTATÁSI EREDMÉNYEK

- kutatási eredmények előállítására vonatkozó előírások
- adatvédelmi szabályok, tájékoztató

1. BEVEZETÉS

Jelen fejezet célja, hogy a KSH Kutatószobájában előállt kutatási eredmények KSH által végzett adatvédelmi vizsgálatának szerepe és működése átlátható, és a lehető legrövidebb átfutási idejű legyen a kutatószobai szolgáltatást használó kutatók számára.

A statisztikai törvény (2016. évi CLV. törvény a hivatalos statisztikáról) szerint

„A KSH (...) tudományos célból hozzáférést biztosíthat olyan egyedi adatokhoz, amelyek megismerése révén a statisztikai egységek közvetlenül nem azonosíthatók.

(...) A biztonságos környezetből csak olyan kutatási eredmény vihető ki, amelyet a Hivatalos Statisztikai Szolgálat tagja az eredmény kiadását megelőzően megfelelő módszerek alkalmazásával ellenőrzött annak érdekében, hogy azokból a statisztikai egységek felfedésének kockázata a mindenkori legjobb statisztikai módszertani megoldásokkal összhangban minimális legyen.”

A KSH tehát a Kutatószobában hozzáférést nyújthat közvetlen azonosítóktól (pl. név, törzsszám) megfosztott adatokhoz, azonban a kutatószobai környezetben előállított kutatási eredményeket, a kutatónak történő kiadását megelőzően, adatvédelmi szempontból minden esetben ellenőriznie, és szükség esetén további adatvédelmi akciók alkalmazásával védenie kell.

Az adatvédelmi ellenőrzés (output checking) célja, hogy a kutatási eredményekből egyedi adatok (pl. egy adott természetes személyhez vagy vállalathoz köthető adatok) felfedésének kockázata¹ minimális legyen.

Hatékony adatvédelmi ellenőrzés akkor valósulhat meg, ha

- a kutatási eredmények áttekinthetők, megfelelően dokumentáltak;
- nincs szükség adatvédelmi eljárások alkalmazására, tehát a kutatási eredmény változtatás nélkül kiadható, mert a felfedés kockázata minimális.

¹ Felfedési kockázat: annak valószínűsége, hogy a közzétett adatok alapján legalább egy statisztikai egység azonosítható, vagy rá vonatkozóan eddig nem ismert információ kerül nyilvánosságra.

Azonosítás: az az esemény, amikor egy adott statisztikai egységet (különösen: természetes személyt, gazdálkodó szervezetet, egyéb intézményt) egyértelműen felismernek vagy hozzájutnak a statisztikai egység egy- vagy több közvetlen azonosítójához.

Felfedés: a közzétett adatokból egy adott statisztikai egységre vonatkozó, eddig nem ismert információ nyilvánosságra kerülése.

A KSH az output checking tevékenység ellátása alatt a kutatás során tett szakmai megállapításokat, az alkalmazott matematikai-statisztikai megoldások helyességét, a levont következtetések szakmai tartalmát nem vizsgálja; a vizsgálat szigorúan csak az adatvédelmi szempontoknak való megfelelésre vonatkozik.

Ezen tájékoztató azokat a szabályokat foglalja össze, amelyeket a kutatónak követnie kell ahhoz, hogy hatékony adatvédelmi ellenőrzés valósulhasson meg. A szabályok betartása a kutató érdeke is, hiszen a kutatási eredmények ellenőrzése rövidebb idő alatt történhet meg, így a kutató hamarabb juthat hozzá az eredményekhez.

A tájékoztató az alábbiakban röviden tárgyalja

- a kutatás dokumentációjára vonatkozó előírásokat,
- az adatvédelmi ellenőrzéshez előállítandó segédfájlokat és információkat,
- a kutatási eredmények ellenőrzésére vonatkozó adatvédelmi szabályokat.

2. A KUTATÁS DOKUMENTÁCIÓJA

2.1 KUTATÓI MUNKA SORÁN ELVÁRT KÖNYVTÁRSTRUKTÚRA

A kutatáshoz dedikált tárhelyen (*Kimeno*\<Felhasználónév>) az alábbiak szerint kell tárolni az adatvédelmi ellenőrzéshez szükséges állományokat.

kutato17

Kovacs_Istvan

kutato17_Kovacs_Istvan_20141030

Kutatasi_eredmenyek: *foglalkoztatottak.dta*
jegyzetek.docx
valtozoleirasok.xlsx
kutatas.do
eredmenyek_1.xls

Munkafajlok: *adattisztitas.do*
leiro_statisztikak.xlsx
export_panel.dta
ipari_termeles2.dta

Programfajlok: *program_merge.do*
program_regressions.xlsx
magyarazat.docx

Adv_ellenorzeshez: *gyak_foglalkoztatottak.dta*
leg_hozz_foglalkoztatottak.dta

kut_eredm_dok_<Vezetéknév>_<Keresztnév>_<ÉvHónapNap>.doc

<Felhasználónév>: A felhasználó név alatt a kutatáshoz hozzárendelt, sorszámokkal ellátott kutatási felületet értünk (pl. kutato17).

<Vezetéknév>_<Keresztnév>: A kutatási felületen egy saját névvel ellátott mappát kell létrehozni a kutatónak. (pl. Kovacs_Istvan).

<Felhasználónév>_<Vezetéknév>_<Keresztnév>_<ÉvHónapNap>: Amikor a kutató eredményt szeretne kivinni, akkor a saját névvel ellátott mappán belül egy alkönyvtárat hoz létre aznapi dátummal (pl. kutato17_Kovacs_Istvan_20141030). Ezen a mappán belül, egymással egy szinten hozza létre a következő négy mappát:

- *Kutatasi_eredmenyek*
- *Munkafajlok*
- *Programfajlok*
- *Adv_ellenorzeshez*

Továbbá ebbe a mappába kérjük elmenteni a dokumentációs űrlapot, a következő névvel **kut_eredm_dok_<Felhasználónév>_<Vezetéknév>_<Keresztnév>_<ÉvHónapNap>.docx**

1. **Kutatasi_eredmenyek** mappa: tartalmaz minden olyan fájlt, amit a kutató *ki szeretne vinni* (táblázatok, regressziós eredmények, programkódok, szöveges dokumentumok, stb.). Ezek az eredmények adatvédelmi szempontból teljeskörűen ellenőrzésre kerülnek kiadásukat megelőzően.

2. **Munkafajlok** mappa: tartalmaz minden olyan állományt (munkaállomány), amely szükséges volt ahhoz, hogy a '*Kutatasi_eredmenyek*'-ben szereplő állományok előálljanak, de a kutatónak *nem áll szándékában kivinni* őket.

3. **Programfajlok** mappa: tartalmazza a kutatási eredmények előállításához szükséges összes programkódot, és egyéb kapcsolódó programkódokat.

4. **Adv_ellenorzeshez** mappa: tartalmaz minden olyan fájlt, amely segíti az adatvédelmi ellenőrzést. Kérjük a mappába menteni a következőket:

- A kutatási eredményekhez tartozó megfigyelések számát, még akkor is, ha az nem szükséges a kutatáshoz. Pl. Tipikusan ilyen eset, ha arányok, megoszlások, százalékos értékek szerepelnek a kutatási eredményben.
- Amennyiben a kivitelre szánt eredmények között értékösszeg² táblázat is található, akkor az értékösszeg táblázatokhoz tartozó segédtáblázatokat:

² Táblázatos adat: mikroadatokból a bennük szereplő változók alapján előállítható olyan táblázatos formába rendezett adatállomány, amely aggregált adatokat tartalmaz. Kétféle táblázatos adatot különböztetünk meg, a gyakorisági táblákat és az értékösszeg táblákat.

Értékösszeg tábla: olyan táblázatos adat, amelynek celláiban az adott mintának vagy teljes sokaságnak az adott cellát meghatározó tulajdonságokkal rendelkező egyedeihez tartozó valamilyen ismérv értékeinek összege vagy átlaga szerepel.

Gyakorisági tábla: olyan táblázatos adat, amelynek celláiban az adott minta vagy teljes sokaság adott tulajdonságaival rendelkező egyedeinek száma szerepel.

Mikroadat: a mikroadat az adatok azon elsődleges formája, amelyből minden más adat származik, és amelyben tárolják őket.

- *abszolút gyakorisági tábla;*
- *legnagyobb hozzájárulónak az értékösszegeből való százalékos részesedésére vonatkozó tábla.*

- Ábrák esetén, az ábrák alapjául szolgáló adatokat.

A segédtablázatok elkészítésére vonatkozó szempontokat lásd a 2.2 fejezetben.

Segédtablák elnevezései: Kérjük, hogy az eredeti értékösszeg táblázat elnevezésétől annyiban különböztesse meg a fájlneveket, hogy azok a „**gyak_**” illetve „**leg_hozz_**” előtagot tartalmazzák. (lásd. *2. melléklet.*)

2.2 SEGÉDTÁBLÁZATOK ELŐÁLLÍTÁSA

Amennyiben a kivitelre szánt eredmények között értékösszeg táblázat is található, akkor az ellenőrzés érdekében a kutatónak el kell készítenie az adott táblázathoz tartozó segédtablázatokat: az *abszolút gyakorisági* táblát, illetve a *legnagyobb hozzájáruló értékösszegeből való részesedését* (százalékos arányát) cellánként, szintén egy külön fájlban. A segédtablázatok **struktúrája meg kell, hogy egyezzen az eredeti táblázatéval**, ugyanazokat a csoportosító ismérveket kell tartalmaznia. (Példa. lásd. *2. melléklet*) A segédtablázatoknak **ugyanabból a mikroadatból kell származniuk**, mint a kivitelre szánt értékösszeg táblának.

- Az *abszolút gyakoriságok* kiszámítása során mindig azt kell vizsgálni, hogy a kivitelre szánt tábla celláiban szereplő értékösszegekhez hány különböző adatszolgáltató járul hozzá nem hiányzó értékkel. (A nulla értéket felvevő adatszolgáltatókat figyelembe kell venni az adatszolgáltatók számának meghatározásakor. Viszont az üres cellákat figyelmen kívül kell hagyni.)
- A *legnagyobb hozzájáruló értékösszegeből való részesedésének* táblázatát az alábbi képlet szerint kell előállítani:

$$\frac{\max(|x_1|, |x_2|, \dots, |x_n|)}{\sum_{i=1}^n |x_i|} * 100$$

ahol x_i az i . adatszolgáltató hozzájárulása az adott cella értékösszegéhez.

A segédfájlokat („**gyak_**”; „**leg_hozz_**”) az adott napi kutatási eredményekhez tartozó 'Adv_ellenorzes' mappába kell helyezni a kutatónak.

2.3 DOKUMENTÁCIÓ A PROGRAMFÁJLOKBAN

A kutatás dokumentálását (lásd. *1. melléklet*) a programfájlokban belül az alábbiakban felsorolt pontok figyelembevételével kérjük elvégezni:

1. A programfájl elején kérjük feltüntetni, hogy mi az *elemzés célja*, milyen állományokon dolgozott (közvetlen vagy közvetett módon).
2. Az elemzés logikája, *felépítése legyen követhető* (pl. adatállományok összeállítása, leíró jellegű elemzések, analitikus eredmények) a programfájlba illesztett megjegyzésekkel,

magyarázatokkal, illetve a programfájl szakaszainak megfelelő tagolásával (cím, rövid leírás).

3. Az átnevezett és az újonnan képzett változók definiálása a programfájlban szövegesen is, mellyel a kutató minden olyan információt megad, amely a változók tartalmára vonatkozik.
4. Ha a kutató több programfájlt használt a kutatás során, akkor érdemes készíteni egy „vezérlő” programfájlt, amely megfelelő sorrendben hívja meg a többi programfájlt. Egy ilyen programfájl megjegyzésekkel ellátva bonyolultabb esetekben nagyban segítheti az átláthatóságot.
5. **A kikért outputokhoz kérjük minden esetben írjon egy-egy magyarázó mondatot a Dokumentációs űrlapra, ezzel megkönnyíti az output értelmezését, ami gyorsabb ellenőrzést és kiadást eredményez.**

A programfájlokat magyar vagy angol nyelvű dokumentációval szükséges ellátni.

2.4 NEM PROGRAMKÓDDAL ELŐÁLLÍTOTT KUTATÁSI EREDMÉNYEK DOKUMENTÁLÁSA

Ha nem programkóddal történik a kutatási eredmény előállítása, akkor is megfelelő dokumentációt kell biztosítani a kutatónak a beadott kutatási eredményekhez. A kutatási eredmény **előállításának folyamatát olyan mélységben kell leírnia**, hogy annak alapján az adatvédelemmel foglalkozó munkatárs akár újra elő tudja állítani a kutatási eredményt (mint ahogy egy programfájl alapján is meg tudná tenni). Minden általa használt, átnevezett, illetve **képzett változót megfelelően, egyértelműen definiálnia** kell. A változókat a használt szoftver lehetőségei szerint **címkeznie** kell.

2.5 A KUTATÁSI EREDMÉNYEK FORMÁTUMA

Kutatási eredményeiket kérjük az alábbi kiterjesztések valamelyikére átkonvertálni:

- táblázatos és egyéb adatok³, eredmények: xls, xlsx, xml, dta, sav, spv, sas7bdat, csv - a csv-t, ha szükséges benne adatvédelmi módosítás, xls formátumban adjuk ki
- ábrák: ppt, jpg, bmp, png, pdf, gph – szerkeszthető gph ábra csak indokolt esetben, részletes leírással kérhető ki. Az „adv_ellenorzeshez” mappába kérjük a mögöttes gyakoriságok és ábrához felhasznált értékek előállítását is. Az ábrák abban az esetben elfogadhatóak, ha a mögöttük lévő adattábla nem tartalmaz alacsony gyakoriságot, egyedi adatot, vagy nem pontosan leolvasható a változó kategóriájának értéke a tengelyekről.
- jegyzet: doc, docx, txt, pdf – csak szöveges leírást tartalmazzon, ne a futási eredményekből kiemelt táblázatokat
- syntax: sps, do, R, Rmd, sas
- log: smcl, log – csak abban az esetben vihető ki, ha nincs szükség adatvédelemre

³ lásd. 2. lábjegyzet.

Ugyanaz a kutatási eredmény nem vihető ki többféle formátumban, mivel ha adott kutatási eredményen szükséges adatvédelmet végrehajtani, akkor pontosan ugyanazt a védelmi eljárást kell alkalmazni minden fájl típus esetén. Ez összetett táblázatos adatok esetén időigényes folyamat.

2.6 DOKUMENTÁCIÓS ŰRLAP

Ha a dokumentációs sablon kitöltése hiányos, vagy nem történt meg, akkor pótlásra kérjük fel a kutatót. Addig nem kezdődik el az adatvédelmi ellenőrzés, amíg a dokumentációs űrlap hiányos. (lásd. 4. melléklet)

3. KUTATÁSI EREDMÉNYEK UTÓLAGOS ADATVÉDELMI SZEMPONTÚ ELLENŐRZÉSE (OUTPUT CHECKING)

Az alábbiakban bemutatásra kerülnek a kutatási eredmények adatvédelmi szempontú ellenőrzése során alkalmazott adatvédelmi szabályok. Ezzel a célunk az, hogy már a kutatási eredmények előállításánál érvényesüljenek az adatvédelmi szempontok és a kutatók számára is érthető legyen, hogy a kutatószobai működésre megállapított szabályokra miért van szükség. Mielőtt adatvédelmi ellenőrzésre küldené az előállított eredményeket kérjük, hogy tekintse át a fejezet tartalmát. Amennyiben az itt leírt szabályok nem teljesülnek, úgy célszerű újból megvizsgálnia az eredményeket, és mérlegelni, hogy mit szükséges módosítani, átalakítani ahhoz, hogy az output minden, lent említett szabálynak megfeleljen. Ha a beadott eredmények a lenti szempontok figyelembevételével készültek is, attól még a KSH szakértőinek azokat ellenőrizniük kell, de a kutató rövidebb várakozási idő után juthat hozzá a kívánt eredményekhez.

Ha a kutató az eredmények előállításánál nem veszi figyelembe az adatvédelmi szabályokat, akkor a KSH a 3.1 fejezetben leírt módon biztosítja az adatok védelmét. Ha adatvédelem szükséges, akkor az előállított táblázatok átstrukturálására, túl részletes kategóriák összevonására, vagy cellaelnyomásra (azok „...” szimbólummal való helyettesítésére) kerülhet sor.

3.1 KUTATÁSI EREDMÉNYEKRE VONATKOZÓ ADATVÉDELMI SZABÁLYOK

Általános szabályok:

1. Mikroadat, mikroadat részlet, döntő többségében alacsony abszolút gyakoriságú cellaértékeket tartalmazó táblázatos adat kivételének tilalma: Kutatószobából mikroadat, illetve mikroadat részlet nem vihető ki. Szintén nem adható ki olyan

táblázatos adat, mely celláinak döntő többsége háromnál kevesebb adatszolgáltató adatát tartalmazza.

2. Hármas szabály: a hármas szabály alapján egy cella érzékenynek tekinthető, ha az adatszolgáltatók száma háromnál kevesebb⁴.

További szabály értékösszeg táblázatra:

3. (*I, k*) dominancia- szabály: értékösszeg táblázatban az (*I, k*) dominancia szabály⁵ szerint érzékenynek tekintendő a cella, ha az értékhez hozzájáruló adatszolgáltatók közül a legnagyobb részesedésének értéke meghaladja a cella értékének *k* %-át.

Statisztikai modellekre vonatkozó szabályok

4. Modell eredményekre vonatkozó szabályok: A modell eredmények (pl. t-próba, F-próba, χ^2 -próba, regressziós eredmények) kivételének feltétele, hogy a modell megfelelő számú megfigyelésen alapuljon (min. 10), továbbá ne történjen modell futtatás egyetlen vállalatra/személyre még idősoros állomány esetén sem.

Ha a kívinni kívánt kutatási eredmények a fent leírt szabályok valamelyikének nem tesznek eleget, azaz adatvédelemi módszerek alkalmazása szükséges, akkor a következőket teheti a kutató, illetve a KSH szakértő:

- **táblázat átstrukturálása/túl részletes kategóriák összevonása**: túl részletes kategóriák használata miatt nagy számban előforduló *érzékeny*, azaz 1-2 adatszolgáltató értékét tartalmazó cellák számának csökkentése a kategóriák összevonásával. Az átalakítást a kutatónak kell elvégeznie, a szakértő csak javaslatot tehet a táblázat átalakítására. (lásd. 3.3 fejezet illetve 3. melléklet.)
- **cellaelnyomás**: Ha a táblázat átstrukturálása nem megoldható, akkor a védendő cellaértékek valamilyen egyezményes jellel (tipikusan a „...”-tal) helyettesítendőek. E feladatot az adatvédő szakértő hajtja végre. (lásd 3.3 fejezet b); c) illetve 3. melléklet.)

⁴ Lakosságstatisztikai adatokat használó táblázatos adatok esetén nem él automatikusan a hármas szabály. A hármas szabály esetleges enyhítéséről az alábbi szempontokat figyelembe véve döntünk: mennyire részletes a vizsgált területi bontás; mekkora a vizsgált populáció/sokaság mérete; milyen a vizsgálatba bevont változók érzékenysége; a vizsgált változók mennyire alkalmasak megfigyelések (adatszolgáltatók) beazonosítására. Az előbbi szempontokat olyan változók esetén vesszük figyelembe, mint például: *földrajzi bontás* (pl. *megye, település, számlálókörzet, stb.*); *osztályozások* (pl. *FEOR, TESZOR, TEÁOR, ÖVTJ, GFO, stb.*); *életkor*; *nem*; *családi állapot*; *születési ország*; *iskolai végzettség*; *gyerekek száma a háztartásban*; *egy háztartásban élők száma*; *jövedelem*; *gazdasági státusz*; *faji eredet*; *nemzetiség*; *egészségi állapot*; *vallási hovatartozás*. A kutatóknak azt javasoljuk, hogy minden kutatási eredmény leadásakor számítsanak a hármas szabály alkalmazására.

Gazdaságstatisztikai állományokból készített kutatási eredményekre automatikusan a hármas szabályt alkalmazzuk függetlenül attól, hogy a felvétel teljes körű, vagy mintán alapszik. Ennek oka, hogy a vállalatokról sokkal több publikus információ áll rendelkezésre, illetve a vállalkozások sokasága jóval kisebb a lakoságinál. Ezáltal az azonosítás esélye is magasabb, mint a társadalomstatisztikai adatok esetén.

⁵ A *k* paraméter értéke nem feltétlenül állandó, függhet például a kiinduló adatoktól, az előállított eredmények részletességétől is.

3.2 AZ EGYES KUTATÁSI EREDMÉNYTÍPUSOK ESETÉN FIGYELEMBE VETT ADATVÉDELMI SZABÁLYOK

| Kutatási eredmény típusa | Adatvédelmi szabályok sorszámai a 3.1 fejezetben megadott 4 alapeset szerint |
|--|--|
| Gyakorisági táblák | 1;2 |
| Értékösszeg táblák | 1;2;3 |
| Maximum, minimum, kvantilisok | *2 |
| Módusz | 2 |
| Átlag, index, arányok | 2;3 |
| Az eloszlás magasabb momentumai (szórásnégyzet, kovariancia, alakmutatók) | 2 |
| Grafikonok: az adat képi megjelenítései | ** |
| Lineáris regressziós együtthatók | 4 |
| Nemlineáris regressziós együtthatók | 4 |
| Becslésekből származó összegző és tesztstatisztikák (t, F, R ² , χ^2 stb.) | 4 |

*Ha a minimum, maximum értékek egy konkrét megfigyelési egységre vonatkoznak, akkor védendőek. A kiadhatósághoz a kutatás dokumentációjából ki kell derülnie, hogy a kívülről kívánt minimum és maximum értékek nem egy adott megfigyelési egység adatát tartalmazzák.

**Semmilyen megfigyelési egység pontos értéke (pl. maximum) nem jelenhet meg adatfelirattal a grafikonon vagy ahhoz csatoltan, pl. a grafikon háttértáblájaként.

3.3. JAVASLATOK

- A) *Táblázat átstrukturálása.* Javasoljuk, hogy az olyan változók esetén, mint például az *életkor, háztartásban élők száma, gyerekek száma*, bizonyos értékek felett alkalmazzanak felső kódolást (Pl. életkor esetén 85+, gyerekek száma esetén 10+) vagy képezzenek csoportokat (Pl. 10-es korcsoport).
- B) *Többdimenziós táblázatok* esetén, ha egy táblázat három vagy több változóból tevődik össze, melyek közül legalább egy érzékeny vagy azonosító változónak tekinthető, akkor az adatvédelmi beavatkozás valószínűsége megnő, még országos szinten is. Ilyen változók lehetnek pl. *születési ország, születési év, legmagasabb befejezett iskolai végzettség szakának megnevezése; nemzetiség, FEOR*.
- C) *Számos változóérték.* Olyan változók esetén, melyek számos kategóriával rendelkeznek, kérjük, csak azokat a kategóriákat válasszák be a kutatási eredményekbe, amelyek a kutatáshoz szükségesek. (Pl. ha az „állampolgárság” 285 lehetséges értékéből a kutatáshoz elegendő csupán 10 nevesített „állampolgárság”, akkor csak a szűkített körből készüljön a kutatási eredmény.)
- D) *Különbségképzés.* Kérjük, figyeljenek a táblázatok különbségeiből eredő adatvédelmi problémákra. Ha pl. az egyik táblázatban a teljes népességre, a másikban pedig a nem roma lakosságra vonatkozó adatok szerepelnek, akkor a két tábla különbségéből a roma

lakosokra vonatkozó jellemzők állnak elő. Hasonló esetekben kérjük, hogy a különbség táblázatot is állítsák elő.

4. KÜLSŐ ÁLLOMÁNYOK

A kutatónak lehetősége van külső fájlokat (jegyzetek, adatbázisok) beküldeni a Kutatószobába, amit majd a kutatás során a kutatószobai környezetben használhat munkája során. A külső állományok kutatószobai környezetbe való bekerülése nem automatikus; ezek az állományok rendelkezésre bocsátás előtt szintén adatvédelmi ellenőrzésen mennek át. Ebben az esetben az adatvédelmi ellenőrzés annak vizsgálatára szorítkozik, hogy az állományban szerepelnek-e közvetlen azonosítók vagy sem. Amennyiben igen, úgy a közvetlen azonosítók eltávolításra kerülnek (pl. név, cím, telefonszám, törzsszám, stb.).

FIGYELEM!

Minden behozni kívánt állomány esetén meg kell adni:

- az állományban szereplő változók metainformációit (változók neve, pontos tartalma)
- az állomány tartalmának megértésére szolgáló rövid leírást Word formátumban.

1. MELLÉKLET: PÉLDÁK PROGRAMFÁJLOK DOKUMENTÁLÁSÁHOZ

Példa 1: A kutatás bemutatása:

„A kutatás során az 5 fő feletti, Magyarországon bejegyzett feldolgozóipari vállalatok termelékenységi viszonyait elemeztem 2002-es adatok alapján. Ehhez ipari termelés (iparitermeles.dta) és a külkereskedelem (kulker.dta) adatait használtam fel. A programfájl elején adattisztítást és az állományok előkészítését, majd összekapcsolását végeztem el. Ezek után TEÁOR kódok alapján leszűrtem a feldolgozóipari vállalatokat, melyekre lineáris regressziókat futtattam. Ezekben azt vizsgáltam, hogy X, Y, Z magyarázó változók szignifikánsan magyarázzák-e W változót. A modellt különböző (a, b, c) almintákra is lefuttattam, végül u, v, h és az e változókra a szokásos leíró statisztikákat is lekértem.

Példa 2: Az eredmények grafikus ábrázolása

```
1 *****
2 ** (3) Az eredmények grafikus ábrázolása **
3 *****
4
5 /* Az egy főre jutó GDP növekedési üteme és az exportált autóalkatrészek
6    növekedési üteme közötti kapcsolat grafikus ábrázolása */
7 twoway (scatter GDP_ann_growth valcomp_ann_growth, msymbol(square) msize(tiny)
8         mcolor(black) mlabel(partnercode) mlabsize(tiny) mlabcolor(black))
```

Példa 3: Értékösszeg táblához tartozó gyakoriságok és legnagyobb hozzájárulások kiszámítása

```
1 ** Példa gyakoriságok és legnagyobb hozzájárulások kiszámítására**
2 *****
3
4 cd "D:\Kutatói_tájékoztató"
5
6 * 2014-es ipari termelés és értékesítés adatbázis
7 use 2014, clear
8
9 * Összes értékesítés nettó árbevétele és átnevezések
10 egen ERT_ARBEV=rowtotal(IABB004 IABB005)
11 rename M005 megye
12 rename M0581 szakagazat
13
14 * Változók felcímkézése
15 label variable ERT_ARBEV "Értékesítés nettó árbevétele"
16 label variable megye "Megye"
17 label variable szakagazat "Szakágazat"
18
19 preserve
20 * Kutatási eredmény: Összes értékesítés nettó árbevétele
21 collapse (sum) ERT_ARBEV, by(megye szakagazat)
22
23 save kut_eredm_1, replace
24
25 * Adv_ell: Gyakorisági tábla
26 restore
27 preserve
28
29 collapse (count) ERT_ARBEV_gyak=ERT_ARBEV, by(megye szakagazat)
30
31 save gyak_kut_eredm_1, replace
```



```

32
33 * Adv_ell: Legnagyobb hozzájárulások
34 restore
35
36 gen ERT_ARBEV_abs=abs(ERT_ARBEV)
37 collapse (sum) ERT_ARBEV_sum=ERT_ARBEV_abs (max) ERT_ARBEV_max=ERT_ARBEV_abs,
38             by(megye szakagazat)
39
40 gen ERT_ARBEV_lh=(ERT_ARBEV_max/ERT_ARBEV_sum)*100
41
42 drop *_max *_sum
43
44 save leg_hozz_kut_eredm_1, replace

```

2. MELLÉKLET: PÉLDA ÉRTÉKÖSSZEG TÁBLÁK ADATVÉDELMI ELLENŐRZÉSÉRE

Ha a kutatási eredmény értékösszeg tábla, akkor a következő lépéseket kell végrehajtani:

1. táblázat: Eredmény, mely a '*Kutatasi_eredmenyek*' mappában megtalálható

| Vállalatok árbevételének összege régióinként és a nemzetgazdasági ágak szerint 2010-ben (fiktív) | | | | | |
|--|-------------------------------|-------------------|-----------------|-------------------|-------------------|
| Régiók | Vizsgált nemzetgazdasági ágak | | | | |
| | A | B | C | D | Összesen |
| 1 | 7 767 971 328 | 211 091 899 472 | 9 943 678 279 | 303 314 418 304 | 532 117 967 383 |
| 2 | 293 999 322 944 | 482 392 636 132 | 195 788 826 974 | 132 201 948 800 | 1 104 382 734 850 |
| 3 | 109 496 225 408 | 8 703 476 032 | 125 583 012 384 | 251 129 981 347 | 494 912 695 171 |
| 4 | 199 566 570 752 | 327 763 841 000 | 97 802 072 160 | 12 144 741 376 | 637 277 225 288 |
| 5 | 275 043 249 600 | 70 936 308 800 | 161 725 637 616 | 430 395 294 040 | 938 100 490 056 |
| 6 | 165 900 728 448 | 126 406 154 216 | 334 304 238 378 | 212 415 489 584 | 839 026 610 626 |
| 7 | 233 459 129 720 | 156 755 016 340 | 8 416 344 064 | 70 604 533 024 | 469 235 023 148 |
| Összesen | 1 285 233 198 200 | 1 384 049 331 992 | 933 563 809 855 | 1 412 206 406 475 | 5 015 052 746 522 |

1. lépés: A kutató elkészíti és elmenti az '*Adv_ellenorzeshez*' mappába a 1. táblázathoz tartozó abszolút gyakorisági táblát. (2. táblázat)

A gyakoriságok kiszámításakor az adott értékösszeghez hozzájáruló különböző, nem hiányzó értékkel rendelkező adatszolgáltatók számát kell figyelembe venni. A segéd táblát mindig az adott értékösszeg táblázat struktúrája szerint kérjük elkészíteni: Pl. ha a táblában a vállalatok TEÁOR nemzetgazdasági ág szerinti bontásban kiszámított árbevételei szerepelnek, akkor a gyakoriság ezekhez az árbevételekhez tartozó különböző, nem hiányzó értékekkel rendelkező vállalatok száma lesz.

Ha egy értékösszeg táblában több különböző változóra is szerepel értékösszeg (pl. árbevétel, profit, hozzáadott érték, stb.), úgy mindegyikhez külön-külön el kell készíteni az abszolút gyakoriságokat (mert pl. eltérő lehet változónként a hiányzó értékek száma).

2. táblázat: Az 1. táblázathoz tartozó abszolút gyakorisági tábla

| Vállalatok száma régióként és a nemzetgazdasági ágak szerint 2010-ben (fiktív) | | | | | |
|--|-------------------------------|-----|-----|-----|----------|
| Régiók | Vizsgált nemzetgazdasági ágak | | | | |
| | A | B | C | D | Összesen |
| 1 | 1 | 42 | 10 | 54 | 107 |
| 2 | 51 | 97 | 40 | 26 | 214 |
| 3 | 19 | 2 | 30 | 57 | 108 |
| 4 | 35 | 69 | 19 | 2 | 125 |
| 5 | 51 | 15 | 34 | 79 | 179 |
| 6 | 33 | 28 | 70 | 43 | 174 |
| 7 | 43 | 35 | 1 | 14 | 93 |
| Összesen | 233 | 288 | 204 | 275 | 1000 |

3. táblázat: Az 1. táblázathoz tartozó, legnagyobb mértékben hozzájáruló részesedésére (%) vonatkozó tábla

| Az értékösszeghez legnagyobb mértékben hozzájáruló vállalat részesedése (%) | | | | | |
|---|-------------------------------|-------|--------|-------|----------|
| Régiók | Vizsgált nemzetgazdasági ágak | | | | |
| | A | B | C | D | Összesen |
| 1 | 100.00 | 4.73 | 95.30 | 3.07 | 1.88 |
| 2 | 3.38 | 2.06 | 4.51 | 7.36 | 0.90 |
| 3 | 9.05 | 91.52 | 7.19 | 3.88 | 2.00 |
| 4 | 4.97 | 3.01 | 10.00 | 54.61 | 1.56 |
| 5 | 3.61 | 12.78 | 6.09 | 2.32 | 1.06 |
| 6 | 5.17 | 7.75 | 2.94 | 4.63 | 1.17 |
| 7 | 4.26 | 5.87 | 100.00 | 13.11 | 2.12 |
| Összesen | 0.77 | 0.72 | 1.05 | 0.71 | 0.20 |

2. lépés: A kutató elkészíti és elmenti az 'Adv_ellenorzeshez' mappába a 1. táblázathoz a legnagyobb hozzájárulónak az értékösszegekből való százalékos részesedésére vonatkozó táblát. (3. táblázat)

Hasonlóan, ha több változóra is szerepel értékösszeg, úgy mindegyikhez külön-külön el kell készíteni a legnagyobb hozzájárulók részesedéseit.

3. lépés: A kutató ellenőrzi az adatvédelmi szabályok teljesülését.

A 2. és 3. táblázatban látható, hogy nem minden esetben teljesül a hármasszabály, illetve bizonyos cellaértékeknél megjelennek domináns hozzájárulók (az (1,80)-dominancia-szabályt alkalmazva). A 4. és 5. táblázat a védendő cellákat emeli ki a (a 2. és a 3. táblázatra vonatkozóan).

A kijelölt cellaértékek védelmét biztosítani kell, ezért a kutató átstrukturálhatja a táblázatot, vagy változtatás nélkül hagyja a kutatási eredményét, elfogadván azt, hogy az eredménytáblájában a kijelölt cellák értékei „...” szimbólummal lehetnek helyettesítve.

| Vállalatok száma régióként és a nemzetgazdasági ágak szerint 2010-ben (fiktív) | | | | | |
|--|-------------------------------|----------|----------|----------|----------|
| Régiók | Vizsgált nemzetgazdasági ágak | | | | |
| | A | B | C | D | Összesen |
| 1 | <u>1</u> | 42 | 10 | 54 | 107 |
| 2 | 51 | 97 | 40 | 26 | 214 |
| 3 | 19 | <u>2</u> | 30 | 57 | 108 |
| 4 | 35 | 69 | 19 | <u>2</u> | 125 |
| 5 | 51 | 15 | 34 | 79 | 179 |
| 6 | 33 | 28 | 70 | 43 | 174 |
| 7 | 43 | 35 | <u>1</u> | 14 | 93 |
| Összesen | 233 | 288 | 204 | 275 | 1000 |

| Az értékösszeghez legnagyobb mértékben hozzájáruló vállalat részesedése (%) | | | | | |
|---|-------------------------------|-------------|--------------|-------|----------|
| Régiók | Vizsgált nemzetgazdasági ágak | | | | |
| | A | B | C | D | Összesen |
| 1 | 100.00 | 4.73 | 95.30 | 3.07 | 1.88 |
| 2 | 3.38 | 2.06 | 4.51 | 7.36 | 0.90 |
| 3 | 9.05 | <u>91.5</u> | 7.19 | 3.88 | 2.00 |
| 4 | 4.97 | 3.01 | 10.00 | 54.61 | 1.56 |
| 5 | 3.61 | 12.7 | 6.09 | 2.32 | 1.06 |
| 6 | 5.17 | 7.75 | 2.94 | 4.63 | 1.17 |
| 7 | 4.26 | 5.87 | <u>100.0</u> | 13.11 | 2.12 |
| Összesen | 0.77 | 0.72 | 1.05 | 0.71 | 0.20 |

4. táblázat (balra fent): Abszolút gyakorisági tábla vizsgálata az adatvédelemhez

5. táblázat (jobbra fent): Legnagyobb mértékben hozzájárulók táblájának vizsgálata az adatvédelemhez

4. lépés: Az adatvédelmi szakértő elvégzi az adatvédelmi ellenőrzést

Tegyük fel, hogy a szakértőnek:

- a 2. táblázat esetében a hármas szabály miatt cellaelnyomást kell végrehajtania (azaz a 4. táblázatban *világos szürke* cellák esetében).
- a 3. táblázat esetében (1, 80) dominancia szabály miatt a 80%-nál nagyobb hozzájárulóknál, cellaelnyomást kell végrehajtania (azaz az 5. táblázat *sötét szürke* cellái esetében).

Ha a 4 és 5. táblázatot összevetjük, akkor látható, hogy összesen 5 cellaértéket kell elnyomni, a fent említett két szabály miatt. Ezt a 6. táblázatban a '...' jel mutatja.

6. táblázat: Cellaelnyomást tartalmazó táblázat

| Vállalatok árbevételének összege régióként és a nemzetgazdasági ágak szerint 2010-ben (fiktív) | | | | | |
|--|-------------------------------|-------------------|-----------------|-------------------|-------------------|
| Régiók | Vizsgált nemzetgazdasági ágak | | | | |
| | A | B | C | D | Összesen |
| 1 | ... | 211 091 899 472 | ... | 303 314 418 304 | 532 117 967 383 |
| 2 | 293 999 322 944 | 482 392 636 132 | 195 788 826 974 | 132 201 948 800 | 1 104 382 734 850 |
| 3 | 109 496 225 408 | ... | 125 583 012 384 | 251 129 981 347 | 494 912 695 171 |
| 4 | 199 566 570 752 | 327 763 841 000 | 97 802 072 160 | ... | 637 277 225 288 |
| 5 | 275 043 249 600 | 70 936 308 800 | 161 725 637 616 | 430 395 294 040 | 938 100 490 056 |
| 6 | 165 900 728 448 | 126 406 154 216 | 334 304 238 378 | 212 415 489 584 | 839 026 610 626 |
| 7 | 233 459 129 720 | 156 755 016 340 | ... | 70 604 533 024 | 469 235 023 148 |
| Összesen | 1 285 233 198 200 | 1 384 049 331 992 | 933 563 809 855 | 1 412 206 406 475 | 5 015 052 746 522 |

Ha tovább vizsgáljuk a táblát észrevehető, hogy a fenti táblázat még *nem védett*, mert például az A1-es cella értéke visszszámolható az *Összesen*-ből. Így további cellákat kell elnyomni, amit másodlagos cellaelnyomásnak nevezünk. Erre több megoldás létezik, egy lehetséges megoldást a 7. táblázatban láthatunk. Így megkaptuk az adatvédelmileg ellenőrzött, kiadható eredményt.

7. táblázat: Kiadható eredmény

| Vállalatok árbevételének összege régióként és a nemzetgazdasági ágak szerint 2010-ben (fiktív) | | | | | |
|--|-------------------------------|-------------------|-----------------|-------------------|-------------------|
| Régiók | Vizsgált nemzetgazdasági ágak | | | | |
| | A | B | C | D | Összesen |
| 1 | ... | 211 091 899 472 | ... | 303 314 418 304 | 532 117 967 383 |
| 2 | 293 999 322 944 | 482 392 636 132 | 195 788 826 974 | 132 201 948 800 | 1 104 382 734 850 |
| 3 | 109 496 225 408 | ... | 125 583 012 384 | ... | 494 912 695 171 |
| 4 | 199 566 570 752 | ... | 97 802 072 160 | ... | 637 277 225 288 |
| 5 | 275 043 249 600 | 70 936 308 800 | 161 725 637 616 | 430 395 294 040 | 938 100 490 056 |
| 6 | 165 900 728 448 | 126 406 154 216 | 334 304 238 378 | 212 415 489 584 | 839 026 610 626 |
| 7 | ... | 156 755 016 340 | ... | 70 604 533 024 | 469 235 023 148 |
| Összesen | 1 285 233 198 200 | 1 384 049 331 992 | 933 563 809 855 | 1 412 206 406 475 | 5 015 052 746 522 |

Ha nem ismertek a sor/oszlop összesenek és nem is áll fenn annak reális lehetősége, hogy egyéb adatforrásokból hozzá lehet jutni, vagy ki lehet számítani, akkor a másodlagos cellaelnyomás alkalmazása nem indokolt.

5. lépés: A kutató megkapja a kiadható eredményt (7. táblázat).

3. MELLÉKLET: PÉLDA TÁBLÁZATOS ADAT ÁTSZUKTURÁLÁSÁRA

Tegyük fel, hogy a kutató az alábbi táblázatot készíti. A gyerekszámot tekintve a táblázat részletes bontást tartalmaz. Amikor a kutató ellenőrzi az adatvédelmi szabályok teljesülését, látja, hogy bizonyos cellaértékeknél nem teljesül a hármas szabály. Ebben az esetben a kutató már tudja, hogy az alacsony cellaértékek védelmét biztosítani kell, tehát átstrukturálja a táblázatot.

8. táblázat: Kutatási eredmény

| Családi állapot | Élve született gyermekek száma Zala megyében | | | | | | | | | |
|-----------------|--|----------|----------|----------|----------|----------|----------|----------|----------|-----------|
| | 1 gyerek | 2 gyerek | 3 gyerek | 4 gyerek | 5 gyerek | 6 gyerek | 7 gyerek | 8 gyerek | 9 gyerek | 10 gyerek |
| hajadon | 17 | 3 | 3 | 3 | | 1 | | | | |
| házas | 54 | 27 | 15 | 11 | 5 | 1 | 1 | 2 | | |
| özvegy | 71 | 28 | 14 | 9 | 1 | 2 | | | 1 | 1 |
| elvált | 11 | 4 | 4 | 5 | | | | | | |

A 8. táblázat átstrukturálása a 9. táblázatban látható. A gyermekek számának felső kódolásával adatvédelmileg megfelelő táblázatot kapunk. A kapott 9. táblázatot továbbítja a kutató adatvédelmi ellenőrzésre.

9. táblázat: Átstrukturált táblázat

| Családi állapot | Élve született gyermekek száma, Zala megye | | | |
|-----------------|--|----------|----------|------------|
| | 1 gyerek | 2 gyerek | 3 gyerek | 4 + gyerek |
| hajadon | 17 | 3 | 3 | 4 |
| házas | 54 | 27 | 15 | 20 |
| özvegy | 71 | 28 | 14 | 14 |
| elvált | 11 | 4 | 4 | 5 |

4. MELLÉKLET: DOKUMENTÁCIÓS ŰRLAP

Kutatási eredmények dokumentációs űrlapja

1. Kutatási alapinformációk

| | |
|--|--|
| 1.1 Kutató bejelentkezési azonosítója: | |
| 1.2 Kutatási projekt megnevezése: | |
| 1.3 Kutatási intézmény neve (szervezet neve): | |
| 1.4 Kutatási eredményt előállító kutató neve: | |
| 1.5 Kutatási eredményt előállító kutató e-mail címe: | |
| 1.6 Kutatási eredmény leadásának dátuma (éééé/hh/nn): | |

A KSH biztonságos környezetéből a kutatási eredmények kivételének szükséges feltétele a kutatási eredmények dokumentációjának helyes kitöltése, melyet a tájékoztatási feladatokat ellátó szervezeti egység munkatársai vizsgálnak át, probléma vagy helytelen kitöltés esetén felveszik a kapcsolatot a kutatóval. **Csak helyesen és hiánytalanul kitöltött űrlapok kerülnek továbbításra a kutatási eredmények utólagos adatvédelmi szempontú ellenőrzését végző szakértők számára.**

2. Kutatási eredményre vonatkozó dokumentáció

Kötelező a válaszadás, kérjük figyelmesen olvassa el a táblázat kitöltésére vonatkozó segédletet.

Vegye figyelembe, hogy a kutatási eredmények fájl formátuma csak az alábbi lehet:

- táblázatos és egyéb adatok, eredmények: xls, xlsx, xml, dta, sav, spv, sas7bdat, csv – a csv-t, ha szükséges benne adatvédelmi módosítás, xls formátumban adjuk ki
- ábrák: ppt, jpg, bmp, png, pdf, gph - szerkeszthető gph ábra csak indokolt esetben, részletes leírással kérhető ki. Az „adv_ellenorzeshez” mappába kérjük a mögöttes gyakoriságokat és ábrához felhasznált értékek előállítását is. Az ábrák abban az esetben elfogadhatóak, ha a mögöttük lévő adattábla nem tartalmaz alacsony gyakoriságot, egyedi adatot, vagy nem pontosan leolvasható a változó kategóriájának értéke a tengelyekről.
- jegyzet: doc, docx, txt, pdf – csak saját gondolatokat tartalmazzon, ne a futási eredményekből kiemelt táblázatokat
- syntax: sps, do, R, Rmd, sas
- log: smcl, log – csak abban az esetben vihető ki, ha nincs szükség adatvédelemre

Kérjük ellenőrizze, hogy elkészítette és bemásolta-e az „Adv_ellenorzeshez” mappába a kutatási eredményeihez tartozó, ellenőrzéshez szükséges segéd táblázatokat: gyakorisági táblákat az értékösszeg táblákhoz és az ábrákhoz, valamint a legnagyobb hozzájáruló részesedésére vonatkozó táblázatokat. (lásd. Kutatási eredmények 2.2. fejezet)

Kitöltési segédlet:

Kutatási eredmény megnevezése: Kérjük tüntessen fel minden eredményt kiterjesztéssel együtt, amit a „Kutatási eredmények” mappába másolt.

Eredmény típusa/előállítási módszere: Például gyakorisági tábla, értékösszeg tábla, regresszió analízis, log fájl. Ha a kutatási eredmény log fájl, akkor nem kell az abban szereplő minden táblát, regressziós outputot stb. egyenként felsorolni, de a programfájlban kérjük ezeket rövid kommentekkel ellátni, illetve egy rövid leírást készíteni róla. **Log fájl csak abban az esetben adható ki, ha nem szükséges adatvédelem!**

Programfájl neve és sorai: Adja meg a programfájl nevét és azon sorait, amelyek a kutatási eredményt előállítják. **Ha programkód nélkül** állította elő az eredményt, akkor kérjük, hogy ezt az oszlopot hagyja üresen, de a 1.2 pontnál röviden ismertesse a kutatási eredmény előállításának célját, a kutatási eredmény előállításának logikai menetét, előállításának főbb lépéseit és adja meg az újonnan képzett vagy átnevezett változók definícióját a 2. táblázatban.

Eredmény előállítása során használt inputállományok (évvel): KSH-inputok esetében pl: Népszámlálás (2011), Halálozás (2009), K+F adatok (2017-19) stb. Külső, a kutató által behozott állományok esetén is szükséges feltüntetni azok pontos megnevezését és ha nem egyértelmű, akkor az időszakot is.

Tartalmi magyarázat és összefüggések: Kérjük adjon egy-egy magyarázó mondatot a kutatási eredmény tartalmáról pl: „Magzati veszteségi ráták országos szint”, „OLS regressziók magas végzettségű vs. alacsony végzettségű nőkre számolt magzati veszteségi rátákkal”, „az adatfájl az iparági klaszterek méretéről és a külföldi vállalatok arányáról tartalmaz leíró adatokat”.

Illetve jelezze, hogy az adott kutatási eredmény összefügg-e jelenlegi vagy korábbi outputjával pl: Korábban település szintű adatokat vizsgált/vitt ki, most megyei szintűt. Korábban csak a 250 fő fölötti vállalatokon végzett elemzést, most a teljes vállalati körön.

1.1 Programkóddal előállított kutatási eredmények

1. táblázat

| Kutatási eredmény megnevezése | Eredmény típusa/ előállítási módszere | Programfájl neve és sorai Programkód nélküli előállítás esetén 1.2 fejezet kitöltése kötelező. | Eredmény előállítása során felhasznált inputállományok | Tartalmi magyarázat és összefüggések |
|--------------------------------------|--|--|---|---|
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

1.2 Programkód nélkül előállított kutatási eredmények

Kérjük, hogy röviden ismertesse a kutatási eredmény előállításának **célját**, a kutatási eredmény előállításának **logikai menetét**, előállításának **főbb lépéseit**!

Kérjük adja meg a programkód nélkül újonnan képzett vagy átnevezett változók **definícióját**!

2. táblázat

| Változó neve | Változó jelentése | Melyik állományban szerepel |
|---------------------|--------------------------|------------------------------------|
| | | |
| | | |
| | | |
| | | |